ELSEVIER

Available online at www.sciencedirect.com

ScienceDirect

Procedia
Social and Behavioral Sciences

Procedia - Social and Behavioral Sciences 103 (2013) 473 – 484

13th International Educational Technology Conference

Improving Turkish language training materials: Grapheme-to-phoneme conversion for adding phonemic transcription into dictionary entries and course books

Özgün KOŞANER[*a], Çağdas Can BİRANT[b], Özlem AKTAŞ[b]

[a]Dokuz Eylul University, Faculty of Letters, Department of Linguistics, Izmir 35260, Turkey
[a]Dokuz Eylul University, Engineering Faculty, Computer Engineering Department, Izmir 35260, Turkey

Abstract

Course materials for teaching Turkish as a second language lack information about the pronunciation of Turkish. These materials have chapters about the alphabet and the pronunciations of the letters, but they lack notation as phonemic transcription. Also Turkish dictionaries lack phonemic transcription information with the false belief that Turkish alphabet is a phonemic one. However there is not a one-to-one correspondence between the letters in the alphabet and the phonemes of Turkish. In this respect, this study aims at developing a grapheme-to-phoneme/allophone conversion tool for Turkish. The output of the software can be used in course materials for teaching Turkish as a second language, and Turkish dictionaries for learners.

Selection and peer-review under responsibility of The Association of Science, Education and Technology-TASET, Sakarya Universitesi, Turkey.

Keywords: dictionary; course materials; phonemic transcription; grapheme-to-phoneme conversion

1. Introduction

Turkish language has become very popular due to Turkey's great potential as a global actor and as a gate to newly emerging markets in the Central Asia. The US government has included Turkish in the mission-critical languages list of the CIA (CIA Values Language Capabilities Among Employees, 2009) and the US universities (Boston University, Texas A&M University, University of Pittsburgh, etc.) have started new programs on Turkish history, culture and history or improved their existing ones. These recent developments caused Turkish to be a promising language for the future and many students from all over the world started to learn Turkish both

[*] Corresponding author ozgunkosaner@gmail.com

Selection and peer-review under responsibility of The Association of Science, Education and Technology-TASET, Sakarya Universitesi, Turkey.
doi:10.1016/j.sbspro.2013.10.363

in their own countries and in the government and private institutions in Turkey. To meet the need for learning Turkish as a foreign language, many universities in Turkey have started graduate programs on Teaching Turkish as a Foreign Language (Dokuz Eylul University, Istanbul University, Hacettepe University, Yildiz Teknik University, etc.). However, the learning materials, especially Turkish dictionaries lack one of the most important information about a language, the pronunciation. Turkish course materials have chapters about the alphabet and the pronunciations of the letters but they lack notation as phonemic transcription. Also the Turkish dictionaries lack phonemic transcription information with the false belief that Turkish alphabet is a phonemic one. However there is not a one-to-one correspondence between the letters in the alphabet and the phonemes of Turkish. In this respect, this study aims at developing a grapheme-to-phoneme/allophone conversion tool for Turkish to resolve the problem mentioned above. The output of this software will be directly accessible by the user and will be intelligible for any language teacher familiar with the phonetic symbols used in IPA (International Phonetic Association) alphabet.

The study is organized as follows: first some basic definitions of the concepts frequently used in the study and the linguistic data which is used as the basis of the study are given. Later a brief review of the literature on grapheme-to-phoneme/allophone conversion is presented. Following the literature review the algorithm of the software developed in this study is discussed in detail. At the last section the output and overall performance of the software is presented with examples.

2. Basic Definitions and Turkish Linguistic Data

In order to explain the grapheme-to-phoneme/allophone conversion process some basic concepts to be used in the study are presented to explain what we intend in using these concepts.

2.1. Graphemes and Turkish writing system

A grapheme is the minimal contrastive unit in the writing system of a language, which may be realized in several forms and usually enclosed in angle brackets (Crystal, 2003). For instance the grapheme < a > may be realized as *A,* a, or *a.* Turkish alphabet have 29 letters; however Turkish writing system comprises of 32 graphemes, 29 letter forms represented in the Turkish alphabet and the graphemes < â >, < î > and < û >, formed by adding the circumflex (^) diacritic on graphemes <a>, <i> and <u> (TDK Yazim Kilavuzu, 2013).

2.2. Phoneme

A phoneme is the smallest distinct sound unit in a language (Matthews 1997) capable of distinguishing meaning. For instance phonemes /s/ and /ʃ/ distinguish the words sap and şap. Various sources (Ergenç, 2002; Özsoy, 2004; Göksel & Kerslake, 2005) present different classifications and therefore different numbers of phonemes for Turkish. In this study we manage a unified approach and combine the data from all of these resources. In our classification there are 32 phonemes in Turkish sound system. The equity between the number of graphemes and phonemes should not misguide us to the conclusion that Turkish has one-to-one correspondence between graphemes and phonemes, since some graphemes may represent more than one phoneme. The list of Turkish phonemes is presented in the table below, with the corresponding graphemes and examples.

Table 1. Phonemes in Turkish

Graphemes		Phonemes
Uppercase	Lowercase	
A	a	ɑ / a
B	b	b
C	c	ʤ
Ç	ç	ʧ
D	d	d
E	e	e
F	f	f
G	g	g / ɟ
Ğ	ğ	*
H	h	h
I	ı	ɯ
İ	i	i
J	j	ʒ
K	k	k / c
L	l	ł / l
M	m	m
N	n	n
O	o	o
Ö	ö	ø
P	p	p
R	r	r
S	s	s
Ş	ş	ʃ
T	t	t
U	u	u
Ü	ü	y
V	v	v
Y	y	j
Z	z	z
Â	â	*
Î	î	*

Û	û	*

* These graphemes do not have phonemic correspondences; however they may cause phonological events such as lengthening the vowel.

As it can be seen in Table 1, Turkish phonemes and graphemes do not always present one-to-one correspondence. Some graphemes represent more than one phoneme, and some graphemes do not have phonemic correspondence, they may have allophonic correspondence which will be discussed in following sections.

2.3. Allophone

An allophone is an audibly distinct variant of a phoneme (Matthews, 1997). Allophones are related sounds which are derived from the same phoneme. For instance Turkish phoneme /n/ has three variants, in other words three allophones [n], [ŋ] and [ɲ]. These variants are used with regard to the phonetic environments they occur as below:

<sen> [sɛ̃n]
<tank> [tʰɑ̃ŋk]
<cenk> [ʤẽɲc]

This allophonic variation is represented as allophonic rule, which covers all allophones of a phoneme and their conditions for occurrence. The allophonic rule for the allophone of /n/ is as follows:

/n/ → [n] / +_____, _____+
[ŋ] / V[back] _____ C[velar, plosive]
[ɲ] / V[front] _____ C[palatal, plosive]

The rule above states that phoneme /n/ transforms into (→) allophones ([n], [ŋ] and [ɲ]) in the phonetic environment (/) stated. Accordingly;

- allophone [n] occurs at the syllable boundaries (+), namely at the beginning of a syllable and at the end of a syllable, as in *nasıl* and *sen*
- allophone [ŋ] occurs after a back vowel (V) and before a velar plosive consonant (C), as in *tank*,
- allophone [ɲ] occurs a front vowel and before a palatal plosive consonant, as in *cenk.*

As it can be seen from one example, Turkish phonology has many nuances for most of the phonemes. The phonemes and their allophones are presented with their examples in Table 2 below. Due to space limitations the allophonic rules, explained with an example above, are not presented in detail.

Table 2. Allophonic Variation in Turkish

Graphemes		Phonemes	Allophones
Uppercase	Lowercase		
A	a	α	α a α̃ ã

B	b	b	b
C	c	ʤ	ʤ
Ç	ç	ʧ	ʧ
D	d	d	d
E	e	e	e ɛ ẽ eˇ
F	f	f	f ɸ
G	g	g / ɟ	g ɟ
Ğ	ğ	*	*
H	h	h	h x ç
I	ı	ɯ	ɯ ɯ̃ ɯˇ
İ	i	i	i ĩ iˇ
J	j	ʒ	ʒ
K	k	k / c	k k^{h} c c^{h}
L	l	ɫ / l	ɫ l
M	m	m	m ɱ
N	n	n	n ɲ ŋ
O	o	o	o ɔ̃ ɔ
Ö	ö	ø	ø œ̃ œ
P	p	p	p p^{h}
R	r	r	r ɾ r̥
S	s	s	s
Ş	ş	ʃ	ʃ
T	t	t	t t^{h}
U	u	u	u ũ uˇ
Ü	ü	y	y ỹ yˇ
V	v	v	v β
Y	y	j	j
Z	z	z	z
Â	â	*	*
Î	î	*	*
Û	û	*	*

2.4. <ğ> Grapheme issue in Turkish

The grapheme <ğ>, read as "*yumuşak ge"* (soft g), lacks a corresponding consonantal sound in standard Turkish, although it is pronounced as a voiced velar fricative in some dialects (Göksel & Kerslake, 2005: 7). Ergenç (2002) and Göksel & Kerslake (2005) describe the effects of <ğ> in detail. These effects can be summarized as below:

- It lengthens the preceding back vowel in when it is in syllable-final or word-final position. The lengthening effect is represented with the symbol : in the IPA phonetic alphabet: *yağmur* [jɑːmuɽ̥], *dağ* [dɑː];
- It may be pronounced as a palatal glide when the preceding vowel is a front one: *eğlen* [ejlɛ̃n];
- It is inaudible between identical back vowels, lengthening the first vowel: *uğur* [uːɽ̥]; *siğil* [siːl];
- If it is between two vowels and these vowels have different distinctive features, in other words they are not identical, it causes a vowel shift. The vowel shift is represented with the symbol ˑ in the IPA phonetic alphabet: *ağıt* [ɑˑɯt], *öğe* [øˑeˇ].

2.5. Syllable and syllable structure in Turkish

Syllable is the basic unit of speech or pronunciation (Bussmann, 1998: 1155). Studies on G2P and also phonology show that in most cases information about the position of a syllable boundary is necessary to define the proper domains for phonological and phonetic rules (van den Bosch & Daelemans, 1993; Demberg, Schmid & Möhler, 2007). As it mentioned in the literature, in Turkish syllable is mostly the domain for phonological rules and allophonic variation. For instance a vowel preceding a nasal consonant in the same syllable becomes nasalized. However a vowel preceding a nasal consonant does not change into a nasal vowel.

gün [gỹn]
günü [gy.nyˇ]

As it can be seen in the examples above, the vowel /y/ (ü), becomes nasalized when it precedes the consonant in the same syllable; but it same vowel preceding the same consonant, but this time in another syllable (syllable boundary is presented with full stop), does not change into its nasal counterpart. In order to define the phonemes and the allophonic variation more accurately the syllable structure should be taken into consideration. Özsoy (2004) gives a detailed account of syllable in Turkish; and this study used the rules presented in Özsoy. The structure of the syllable in Turkish can be summarized as below (Özsoy, 2004: 97-98, 101):

Turkish has six different syllable patterns:

- V o
- CV bu
- VC aç
- CVC tok
- VCC alt
- CVCC sarp

Consonant clusters comprise of two consonants and are limited to two consonants only: *Türk, *plan,*

Turkish syllables are mostly open syllables. When a word ending in a consonant is added a suffix beginning with a vowel re-syllabification process occurs and the consonant at the end of the first syllable shifts into the next syllable and becomes onset. *aç – ım → a + çım*

Turkish has some borrowing syllable structures for loanwords.

- VCC aks
- CCV pro

- CCVC spor
- CCCVC Strazburg

3. Literature Review

In order to convert the graphemes into phonemic symbols a computer process called the Grapheme-to-Phoneme conversion is used. Grapheme-to-Phoneme (G2P) conversion is a basic module of any text-to-speech system, which converts a written text to synthesized speech signal. G2P is the task of converting a string of graphemes (letters used in writing) to a string of phonetic symbols (letters used to represent speech sounds) (Demberg, Schmid & Möhler, 2007: 96). G2P has application in many areas such as speech recognition, automatic dialogue systems, and in transliteration systems (systems which provide conversion from one alphabet system to another; e.g., from Cyrillic to Roman).

In the literature there are numerous G2P studies for various languages, in which G2P systems are both used as modules in text-to-speech systems and as independent software (Daelemans & van den Bosch, 1993; Kienappel & Kneser, 2001; Demberg, Schmid & Möhler, 2007; Bisani and Ney, 2008). The G2P systems developed in these studies are also used in the transliteration of languages with different orthography.

G2P systems are also present in the national literature (Salor, 1999; Bozkurt, 2000; Sak, 2000; Şayli, 2002; Salor, 2005; Sak, Güngör & Safkan, 2006; Görmez & Görmez, 2008; Yılmaz, 2009; Akbulut, Adıguzel & Yılmaz 2011). However all of these systems are parts of text-to-speech or speech-to-text systems and their outputs are available for the computer systems only and cannot be accessed and used by the user.

There is one valuable resource on Turkish pronunciation "Konuşma Dili ve Türkçenin Söyleyiş Sözlüğü", a pronunciation dictionary by Ergenç (2002). However, this publication is available only in printed medium and it requires great effort to digitize this publication through OCR methods. Therefore it cannot be used easily in course materials and in Turkish dictionaries.

4. Method

When the literature on Grapheme-to-Phoneme conversion is reviewed it can be seen that this task can be performed using either rule (knowledge) based approach or a data-driven approach (Kienappel & Kneser. 2001):

Rule based approach;

- requires great effort and expertise, and explicitly coded, language-specific, linguistic knowledge sources,
- cannot be applied to new tasks and languages easily,
- has inherent problems with maintenance, it is difficult to change some of the rules without introducing unwanted side effects.
- demonstrates impressive performance for some tasks,

On the other hand, data-driven approach;

- requires spelling of words with their associated pronunciation in a phonemic or phonetic alphabet
- is highly flexible
- is automatically produced for a language according to the rules implicit in the training data, without explicit modelling of linguistic knowledge

In rule based approach to the design of grapheme-to-phoneme modules, explicitly coded, language-specific, linguistic knowledge sources are necessary for performing the task. However, great effort and expertise are required to set up such a module and to adapt this module to a new language or task (Kienappel & Kneser, 2001). In data-driven approach which uses supervised learning techniques, based on a corpus of transcribed words, the same even better performance can be achieved, without explicit modelling of linguistic knowledge (van den Bosch & Daelemans, 1993). But modules, developed using this approach, are applied to corpora, thus require annotated data (a database of word-pronunciation pairs) to be at hand already. (Daelemans & van den Bosch, 1993).

In this study we implement the rule-based approach to G2P since there is not any transcribed or machine readable corpus of words in Turkish to enable us to implement a data-driven approach. Only resource for this purpose is the “Konuşma Dili ve Türkçenin Söyleyiş Sözlüğü”, a pronunciation dictionary by Ergenç (2002). However, this publication is available only in printed medium and it requires great effort to digitise this publication through OCR methods. Instead, we intend to start from scratch, building a customisable architecture that will store the inventory of graphemes from Turkish alphabet, phonemes and phonological rules from phonology in a database.

5. Rule-based G2P software for Turkish

The architecture of the software developed is as follows:

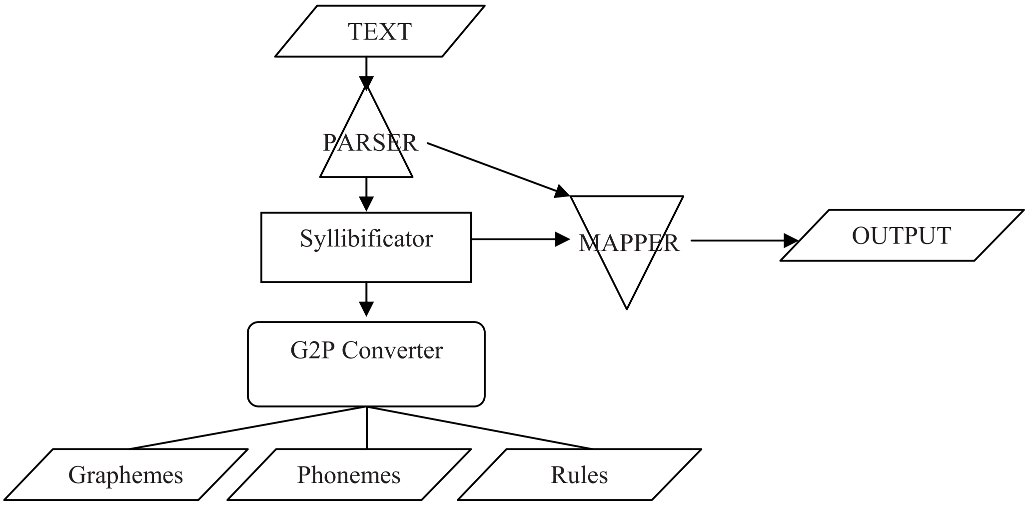

Fig. 1. The architecture of G2P Software

According to this architecture, in the G2P conversion tool for Turkish, the *Parser module* in the first step takes as its input any Turkish text and it parses this body of text into sentences and into words.

In the *syllabificator module*, these parsed spelling of the words are split into their syllables, since in most cases information about the position of a syllable boundary is necessary to define the proper domains for phonological and phonetic rules (van den Bosch & Daelemans, 1993; Demberg, Schmid & Möhler, 2007).

Later in the *G2P converter module* the graphemes in these syllabified words are matched to their phoneme symbols using the phonological rules stored in the inventory.

In the last stage, a *Mapper module* matches the word spelling and phonemic transcription in pairs and produces as its output a corpus of words with their transcription.

5.1. Parser Module

The parser module uses the sentence boundary detection algorithm proposed by Aktaş (2006) and Aktaş & Demir (2006) to parse the plain text files into sentences and words. It stores this annotated data in XML format. This module runs online as a web-service.

5.2. Syllibificator Module

This module splits the parsed words into their syllables using a syllabification algorithm developed using the phonological rules mentioned above.

In order to evaluate the accuracy rate of the syllabificator module 5000 words (two or more syllables) from TDK Turkish Dictionary were randomly chosen. The evaluation results are as below:

Table 3. Accuracy Rate for the Syllibificator Module

Number of Words	5000
Number of incorrectly parsed syllables	22
Number of Correctly Parsed Syllables	4978
Accuracy Rate	*99.56*

As mentioned before syllabification is necessary to define the proper domains for phonological and phonetic rules, and in Turkish phonology in most cases syllable boundaries are important in defining the phonemic environment, namely the context a phoneme occurs.

5.3. G2P Converter

G2P converter module takes the parsed and syllabified words as its input and using the data from the inventory sets for graphemes, phonemes and phonological symbols it converts the graphemes into phonemes according to the phonological rules. This module first maps the graphemes to their corresponding *phonemic* symbol in the alphabet of the IPA. After this first step the module applies the phonological rules on these phonemic symbols and returns the final symbols that are altered according to the allophonic variations represented in the rules. The module makes use of three sets of data in its operations, namely the inventory of graphemes, the inventory of phonemes and the inventory of phonological rules

5.4. Mapper

Mapper module takes the parsed text and the output of the G2P module as its inputs and pairs these data into a corpus of words with their transcription, with and without syllable marking full stops (.).

6. Overall evaluation of the G2P software for Turkish

The software was implemented following the structure mentioned in the previous chapter, after building up the inventory of graphemes and phonemes. Later, the phonological rules were gathered from the resources and encoded in a way that the software could understand these rules.

The software uses UTF-8 Unicode encoding system in presenting the IPA symbols for phonemes due to Unicode's widespread support for IPA symbols and for the ease of use for the end-users who are most probably familiar with the symbols in the IPA.

The software takes any pieces of text as input through an "Input Text" box or from a plain text file encoded in UTF-8 Unicode encoding. Then it converts the graphemes in this "Input Text" into phonemes and displays the results in an "Output Text" box. The user may save the results in a plain text file. The screenshot of the program is given in Figure 2 below.

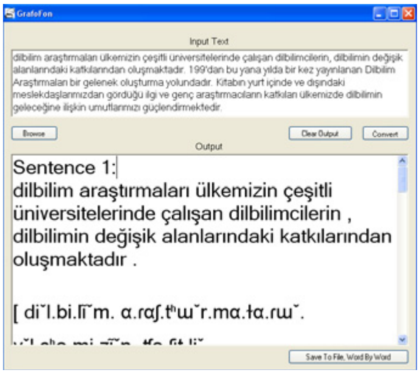

Fig. 2. A screenshot of the G2P software user interface

The software saves the output word by word, with each word's phonemic transcription, both with and without syllable boundaries. This saving method is preferred for future studies; especially a G2P system that uses data-driven approach for the process. Thus, the output of the program is saved in such a structure that any data-driven G2P software can access the grapheme data, the phoneme data and, in addition, the syllable data of the words in the text file. The screenshot of the resulting text file is presented in Figure 3 below.

0. dilbilim - diˇl.bi.lĩˇm. - diˇlbilĩˇm
1. araştırmaları - ɑ.rɑʃ.tʰɯˇr.mɑ.ɫɑ.rɯˇ. - ɑrɑʃtʰɯˇrmɑɫɑrɯˇ
2. ülkemizin - yˇl.cʰe.mi.zĩˇn. - yˇlcʰemizĩˇn
3. çeşitli - ʧe.ʃit.liˇ. - ʧeʃitliˇ
4. üniversitelerinde - y.ni.vɛr.si.tʰe.le.rĩˇn.deˇ. - ynivɛrsitʰelerĩˇndeˇ
5. çalışan - ʧɑ.ɫɯ.ʃɑ̃n. - ʧɑɫɯʃɑ̃n
6. dilbilimcilerin - diˇl.bi.lĩˇm.ʤi.le.rĩˇn. - diˇlbilĩˇmʤilerĩˇn
7. dilbilimin - diˇl.bi.li.mĩˇn. - diˇlbilimĩˇn

Fig. 3. A screenshot of the output file

The software was tested using a 1000 word sample selected from the Turkish dictionary of TDK. The words comprised of two syllables minimum. The 1000 words selected for the sample contained 7894 graphemes. The results of the accuracy evaluation on the software's output are given in Table 4 below.

Table 4. Accuracy rate of the G2P software

Number of Words	1000
Number of Graphemes	7894
Number of Correctly Parsed Graphemes	7588
Number of Incorrectly Parsed Graphemes	306
Accuracy Rate	*96.12*

7. Conclusion

The software implemented in this study is a rule-based grapheme to phoneme system that converts graphemes of Turkish into their corresponding phonemes. The process of implementation comprised of gathering the list of graphemes, phonemes and phonological rules into a database. The resulting software is tested on a word list comprising of the entries of The Turkish Dictionary obtained from TDK. The results of the program are successful and the list of words with their phonetic transcriptions will be presented to TDK to be used in their online dictionaries. The output of the program may be used in text-to-speech and speech-to-text systems which have become popular recently. The output can also be used course materials in Teaching Turkish as a Foreign Language programs or other linguistic research on phonological properties of Turkish Language. The G2P software developed in this study is a rule based G2P system and thus is the first step in grapheme-to-phoneme conversion process for Turkish. The software may give way to other software, for instance data-driven G2P implementations or may be used as an embedded module in text-to-speech systems.

Acknowledgements

This work was supported by Dokuz Eylul University, Project No: 2012.KB.SOS.3

References

Akbulut, A., Adıgüzel, T. & Yilmaz, A. E. (2011). Statistical Syllable Analysis for Pronunciation Ambiguity Detection and Resolution in Text-to-Speech Synthesis Applications: A Case Study in Turkish. *Acta Polytechnica Hungarica, Vol 8, No 5.*

Aktaş, Ö. & Demir, U. (2006). Türkçe için Cümle Sonu Belirleme Yöntemine Yeni bir Bakis. *Paper presented in ASYU 2006-Akilli Sistemlerde Yenilikler ve Uygulamalari Sempozyumu,* Yildiz Technical University, İstanbul.

Aktaş, Ö. (2006). Türkçe için Verimli bir Cümle Sonu Belirleme Yöntemi. *Paper presented in Akademik Bilişim 2006 / Bilgi Teknolojileri Kongresi IV*, Pamukkale University, Denizli.

Birant, Ç. C. (2009). *Türkçe için Heceleyici Yazılım Altyapısının oluşturulması*. Unpublished Manuscript.

Bisani, M. and Ney, H. (2008). "Joint-Sequence Models for Grapheme-to-Phoneme Conversion". Speech Communication, Volume 50, Issue 5, pp. 434-451.

Boston University (2013). Retrieved from: http://www.bu.edu/mlcl/home/why-study-turkish/

Bozkurt, B. (2000). *Reading Aid for Visually Impaired: A Turkish Text-to-speech System Development.* Unpublished MA thesis.

Bussmann, H. (1998). *Dictionary of Language and Linguistics*. London: Routledge.

CIA Values Language Capabilities Among Employees. (2009). Retrieved from: https://www.cia.gov/news-information/featured-story-archive/2010-featured-story-archive/cia-values-language-capabilities.html

Crystal, D. (2003). *A Dictionary of Linguistics and Phonetics.* Malden: Blackwell Publishing.

Daelemans, W. & van den Bosch. (1993). A. TABTALK: Reusability in Data-Oriented Grapheme-to-Phoneme Conversion.Proceedings of Eurospeech 1993, Berlin, pp.1459-1466.

Demberg, V., Schmid, H. & Möhler, G. (2007). Phonological Constraints and Morphological Preprocessing for Grapheme-to-Phoneme Conversion. In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics,* pp. 96-13.

Ergenç, İ. (2002). *Konuşma Dili ve Türkçenin Söyleyiş Sözlüğü.* İstanbul: Multilingual.

Göksel, A.& Kerslake, C. (2005). *Turkish. A Comprehensive Grammar*. London/NY: Routledge.

Görmez, Z. & Orhan, Z. (2008). TTTS: Turkish Text-To-Speech System. *12th WSEAS International Conference on COMPUTERS, Heraklion, Crete Island/Greece, June. 2008, Proceedings of 12th WSEAS International Conference on COMPUTERS*, pp. 977-982.

International Phonetic Association. *http://*www.langsci.ucl.ac.uk/ipa/ 06.08.2012

Kienappel, A. K. & Kneser, R. (2001). Designing very compact decision trees for grapheme-to-phoneme transcription. Paper presented in *7th European Conference on Speech Communication and Technology 2nd INTERSPEECH Event. EUROSPEECH-2001*, pp. 1911-1914.

Matthews, P. H. (1997). *The Concise Oxfor Dictionary of Linguistics*. Oxford: Oxfor University Press.

Özsoy, S. (2004). *Türkçenin Yapısı 1 Sesbilim.* İstanbul: Boğaziçi Üniversitesi Yayınevi.

Sak, H., Güngör, T. & Safkan, Y. (2006). A corpus-based concatenative speech synthesis system for Turkish. Turkish Journal of Electrical Engineering and Computer Sciences, 14(2):209-223.

Salor, Ö. (1999). *Signal Processing Aspects of Text to Speech Synthesizer in Turkish*. Unpublished MA thesis.

Salor, Ö. (2005). *Voice Transformation and Development of Related Speech Analysis Tools for Turkish.* Unpublished PhD thesis.

SAMPA (Speech Assessment Phonetic Alphabet) (2013). Retieved from: http://www.phon.ucl.ac.uk/home/sampa/index.html

Şayli, Ö. (2002). *Duration Analysis and Modelling for Turkish Text-to-Speech Synthesis.* Unpublished MA thesis.

Texas A&M University. (2013). Retrieved from: http://studyturkish.tamu.edu/

Türk Dil Kurumu. (2013). http://www.tdk.gov.tr/index.php?option=com_gts&arama=gts&guid= TDK.GTS.501f72499ea1a8.98681729

University of Pittsburgh. (2013). Retrieved from: http://www.lctl.pitt.edu/turkish.html

Van Bael, C. P. J. (2007). *Validation, Automatic Generation and Use of Broad Phonetic Transcriptions*. Unpublished Dissertation. Radbound Universiteit Nijmegen.

van den Bosch, A. & W. Daelemans (1993). Data-oriented methods for grapheme-to-phoneme conversion. Proceedings of European Chapter of ACL, Utrecht, 45-53.

Yılmaz, A. E. (2009) Türkçe Metinden Konuşma Senztezleme Uygulamaları için bir Veri Sözlük Seti ve Yazılım Çerçevesi. *Gazi Üniversitesi Mühendislik ve Mimarlık Dergisi, Cilt 24, No 4,* pp. 735-744.

Wells, J. C. (1996). Why phonetic transcription is important. In: *Malsori* (Journal of the Phonetic Society of Korea) *31-32*, pp. 239-242.